Supporting Information S1 for 'Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity'

Wilson J. Wright, Kathryn M. Irvine, Emily S. Almberg, and Andrea R. Litt

Alternative model representation

The model described in the main text can also be represented without explicitly including the latent detection counts for each species (Y_{ijk}). This form of the model is useful computationally and for the comparisons to alternative false positive occupancy models (Supporting Information S2). First we show that, given main text equations 2 and 3, the classified counts for a species follow independent Poisson distributions. Dropping some of the subscripts for clarity, let $Y \sim \text{Poisson}(\lambda)$ and $[\mathbf{C} \mid Y = y] \sim \text{Multinomial}(y, \mathbf{\Theta})$ where $\mathbf{C}$ and $\mathbf{\Theta}$ are K-length vectors. The count of detections classified as each species are denoted by $\mathbf{C} = [C_1, \ldots, C_K]$ and the corresponding probabilities of classifying a detection to each species are given by $\mathbf{\Theta} = [\theta_1, \ldots, \theta_K]$. Note that by definition $\sum_{k=1}^{K} C_k = Y$ and $\sum_{k=1}^{K} \theta_k = 1$. We are interested in the marginal distribution for $\mathbf{C}$, but since Y is defined by the summation of $\mathbf{C}$ we can write this as the joint distribution. Consequently, we have

$$
\begin{aligned}
\Pr(\mathbf{C} = \mathbf{c}, Y = y) &= \Pr(\mathbf{C} = \mathbf{c} \mid Y = y)\Pr(Y = y) \\
&= \left[\frac{y!}{c_1! c_2! \cdots c_K!} \theta_1^{c_1} \theta_2^{c_2} \cdots \theta_K^{c_K}\right] \left[\frac{1}{y!} \lambda^y e^{-\lambda}\right] \\
&= \frac{\theta_1^{c_1} \theta_2^{c_2} \cdots \theta_K^{c_K}}{c_1! c_2! \cdots c_K!} \lambda^{c_1} \lambda^{c_2} \cdots \lambda^{c_K} e^{-\lambda\theta_1} e^{-\lambda\theta_2} \cdots e^{-\lambda\theta_K} \\
&= \left[\frac{1}{c_1!} (\lambda\theta_1)^{c_1} e^{-\lambda\theta_1}\right] \left[\frac{1}{c_2!} (\lambda\theta_2)^{c_2} e^{-\lambda\theta_2}\right] \cdots \left[\frac{1}{c_K!} (\lambda\theta_K)^{c_K} e^{-\lambda\theta_K}\right].
\end{aligned}
$$

The probability mass function for each C_k is that of a Poisson distribution with rate parameter $\lambda\theta_k$. Each C_k is independent of each other count because the joint distribution for $\mathbf{C}$ is the product of functions for each individual C_k.

Then the marginal distributions for each $C_{ij\cdot k'}$, the classified detection counts that are observable data, can be modelled directly instead of utilizing the distributions conditional on the latent counts, Y_{ijk}. Under the assumption that all detection counts are independent across species, each $C_{ijkk'}$ variable follows a Poisson distribution and is independent of all other counts of classified detections. Based on the properties of the Poisson distribution, these characteristics mean that each $C_{ij\cdot k'}$ also follows a Poisson distribution with corresponding rate parameter $\sum_{k=1}^{K} z_{ik}\lambda_{ijk}\theta_{kk'}$ given the occupancy states for each species at a site. Consequently, this entire model can be described by main text equation 1 and

$$[C_{ij\cdot k'} \mid \mathbf{Z}_i] \sim \text{Poisson}\left(\sum_{k=1}^{K} z_{ik}\lambda_{ijk}\theta_{kk'}\right),$$

where $\mathbf{Z}_i$ is the vector of occupancy states at site i for all K species.